Interpreting the results of cross-cultural cognitive interviews: a mixed-method approach

Miller, Kristen; Willis, Gordon; Eason, Connie; Moses, Lisa; Canfield, Beth

INTERPRETING THE RESULTS OF CROSS-CULTURAL COGNITIVE INTERVIEWS

A Mixed-Method Approach

*KRISTEN MILLER, GORDON WILLIS, CONNIE EASON, LISA MOSES & BETH CANFIELD**

Cognitive interviewing is used to empirically assess the ways in which individuals mentally process and respond to survey questions that are presented in either written or auditory form, and is commonly applied at the questionnaire pretesting stage (DeMaio & Rothgeb, 1996; Snijkers, 2002; Willis, 2005). However, there is little consensus among practitioners regarding the standards or criteria that constitute high-quality cognitive evaluations (Snijkers, 2003). While some limited research within the evaluation literature pertains to interviewing technique, regarding issues such as specificity versus generality of probes (Foddy, 1998), or concurrent versus retrospective probing (Redline, Smiley, Lee, et al., 1998), few investigations have attended to the analysis of cognitive interview results (Willis, 2005). That is, how are findings from individual cognitive interviews to be used in order to make conclusions regarding the functioning of a survey question? This analytic deficiency has raised skepticism regarding the replicability, falsifiability and, ultimately, the validity of cognitive interview findings (Conrad, Blair & Tracy, 2000; Tucker, 1997; Willis, 2005).

Two strands of thought appear to shape the discussion regarding the quality of information produced by cognitive interviews. The first argues that if the method is to provide meaningful results, cognitive interviews must be standardized; only structured interviews can be systematically analyzed for unbiased results (Tucker, 1997). The second argues that it is the qualitative results of loosely structured interviews – those allowing for spontaneous or emergent probing – that lends strength to cognitive interviewing methodology (Gerber, 1999). Unlike regimented interviews, semi-structured interviews may capture contextual information that is essential to understanding the interpretive aspects of the question-response process.

* The authors would like to thank Jennifer Madans of NCHS for her support and very helpful comments.

Underlying this discussion is the age-old epistemological discourse that results in the pitting of quantitative and qualitative methodologies. It is not our position that only quantitative methods allow for objective and systematic analyses of cognitive interviews. To the contrary, qualitative methodology has produced a vast literature on the systematic analysis of textual data which aptly defends the validity of such an analysis (Denzin & Lincoln, 2000). We suspect, however, that cognitive interviewing methods – while advantaged by the semi-structured interview – can be further enhanced by quantitative analysis, that is, when a mixed-method approach is used (Tashakkori & Teddlie, 1998). A mixed method approach incorporates the contextual data derived from semi-structured cognitive interviews along with numerical coding of results. The combination of qualitative and quantitative analyses presumably augments the quality as well as the variety of information that can be obtained through cognitive interviewing.

Further, we suspect that mixed-method approaches will be especially useful within a vital emerging area: The application of cognitive interviewing techniques across culture and language. Although there are several challenges to quality in cross-cultural cognitive interviewing (Johnson, 1998), the most vexing may be a reliance on a purely qualitative and sometime impressionistic interviewing approach. Because cognitive interviewers typically are nested within cultural or language group (i.e., they can only conduct interviews in languages they can speak, and must therefore employ bilingual staff to conduct other-language interviews), it is not clear whether the results across subgroups represent differences between the cultures represented (or questionnaire translations), or whether they simply reflect stylistic differences between disparate cognitive interviewers. Especially for purposes of assessing the cross-cultural equivalence of survey questions, it is vital that subgroup variation be adequately interpreted, lest investigators be led astray by variance that is only imposed by the question evaluation procedure. Hence, we propose that cross-cultural investigations may benefit from a more structured approach than has typically been employed for purposes of question pretesting, in order to minimize error due only to the interviewer.

Warnecke, Johnson, Chavez, et al. (1997) report on the application of a system for coding cognitive interviewing outcomes, in a study involving racially and culturally diverse populations. However, they provided little information concerning the nature of their coding system, other than implying that it was useful in systematically disentangling the results of a large number of cognitive interviews, and their codes appear to relate mainly to question comprehension processes. For the current study we therefore developed an explicit coding system that emphasizes the full range of cognitive processes postulated to influence survey responding. Codes were based on a cognitively-oriented question-response model of the type described by Tourangeau (1984) depicting (1) the interview

subject's interpretations of key terms, (2) the subject's ability to retrieve information necessary to answer the questions; (3) decision processes used to modify or further process information, or to judge its adequacy; and (4) matching of the respondent's internal representation of their answer to given response categories. These codes were applied to the unstructured written interviewer notes, to produce the quantitative component of a mixed-method analysis of interviews. The paper will describe the findings related to the tested survey questions, using an analytic approach in which the coding and tabulation of results was supplemented by interviewers' open-ended text comments. Further, we examined a range of respondent characteristics other than ethnicity (e.g. gender, ethnicity, age, language) that might impact the question response process. Finally, the paper describes the strengths and weaknesses and methodological utility of this analytic approach.

1 Method

Sample. Sixty-seven cognitive interviews were conducted among (self-reported) Hispanic and Non-Hispanic (both White or Black) participants in urban (Washington D.C.) and rural/suburban locations (two locations in Northwest Ohio). Because of the socio-cultural focus, a relatively equal distribution of participants across socio-economic status, gender, age, ethnicity, as well as language and geographical location, was selected (see Table 1). Participants were recruited through newspaper ads, fliers, and by word-of-mouth. The D.C. area interviews were conducted in the Questionnaire Design Research Laboratory at the National Center for Health Statistics. Ohio interviews were conducted either in the participant's home or in a private room of a community facility. All participants were remunerated $35 after the interview.

Table 1 Northwest Ohio & DC Metropolitan Cognitive Interview Subjects

	Race/Ethnicity	Income	Education	Age	Gender
DC Metropolitan (English) 12 Participants	White = 7 Black = 5	11-20K = 4 21-30K = 1 31-50K = 3 51-80K = 1 61-80K = 0 81K+ = 3	Elementary = 0 Some High School = 2 H. S. Grad. = 4 Some College = 6	18-29 = 0 30-49 = 0 50-69 = 7 70+ = 5	Female = 7 Male = 5
NW Ohio (English) 20 Participants	White = 19 Hispanic=1 Black = 0	0-10K = 7 11-20K = 6 21-30K = 5 Unknown = 2	Elementary = 3 Some H.S. = 7 H. S. Grad. = 8 Some College = 2	18-29 = 1 30-49 = 6 50-69 = 8 70+ = 5	Female = 12 Male = 8
NW Ohio (Spanish, Spanish/English combination) 35 Participants	Mex. Am. = 17 Mexican = 14 Puerto Rican = 1 Hispanic = 1 Hisp. Am. = 1 Cuban = 1	0-10K = 9 11-20K = 9 21-30K = 6 31-50K = 7 51-80K = 4	Elementary = 5 Some H.S. = 12 H. S. Grad. = 13 Some College = 2 Unknown = 3	18-29 = 7 30-49 = 16 50-69 = 11 70+ = 1	Female = 20 Male = 15

Data Collection. The interviews were based on an interviewer-administered health survey questionnaire containing items selected from the NCHS National Health Interview Survey, or created anew, covering chronic conditions, cancer screening, diet, physical activity and demographic characteristics. All but one of the cognitive interviews of Hispanics were conducted in Spanish, and all Non-Hispanics interviews were in English. The instrument was translated from English by one of the authors. The cognitive interviews were semi-structured; along with the survey questions, the interview guide (protocol) consisted of several pre-scripted follow-up questions pertaining to participants' interpretations of key terms and overall comprehension of questions. These fixed probes ensured that this particular information was collected in every interview and could then be compared across all interviews. As a less standardized approach, interviewers were also instructed to inquire as to the ways in which participants constructed their answers to the survey questions, which further provides insight into potential sources of response error. These *emergent*, non-scripted probes helped interviewers make sense of gaps or contradictions in participants' explanations and provided contextual information needed to precisely define question problems. In turn, this open-ended information contributed to the development of a coding system for purposes of succinctly characterizing the results in quantifiable form.

Code Development. Two sets of numerical codes were developed from the cognitive interviews, a *problem set* and an *interpretive set*. The problem codes, based on the standard question-response model (Comprehension, Retrieval, Decision and Response), indicate situations in which tested subjects deviated from or were unable to fully negotiate stages of the question response process. Table 2 outlines this set of codes. It should be noted that problem codes do not necessarily reflect the presence of actual response error. For example, many participants were unfamiliar with the term *chronic obstructive pulmonary disease*, yet based on the total evidence obtained, appeared to respond correctly that they did not have that condition. Nevertheless, because these participants were unable to fully comprehend the question, a problem code of 1 was assigned in such a case.

Because problem codes were based on the response process model, most of the codes were developed prior to interviews. However, a few response problems were not anticipated in the initial schema (e.g. codes 5 and 7) and were added as interviews were being conducted. Consequently, refinement of codes occurred inductively, which ensured that all nuances of question problems were included within the coding schema. Though this process established a complete data set, as a new code was developed, all previous interviews needed to be re-checked for consistency – an extremely time-consuming endeavor.

Table 2 Question-Response Problem Codes

Comprehension	1	*Term*: Subject does not understand or know the meaning of specific words
	2	*Question*: Subject does not understand the question as a whole because of vagueness or complexity
Retrieval	3	Subject does not know (and never knew) the requested information
	4	Subject is unable to remember requested information
Decision	5	Subject is unable to make calculations necessary to arrive at the answer
	6	Question sensitivity or perceived negative reaction by subject
	7	Subject is unable to decide on a response
	8	Subject is found to estimate either too high or too low
Response	9	Response categories do not match subject's internal representation of the answer
----	0	No problems observed

In addition to response problems, codes were developed based on interpretive patterns. These codes were not necessarily directly error-related, but reflected variation in the ways participants conceptualized key terms, such as *health, mid-day meal, advice*, and *exercise*. Unlike the problem codes, these interpretive codes were generated entirely from an inductive process, that is, based on qualitative analysis of the interview text. After interviews were collected, patterns of interpretation were identified across participants, and each pattern was then assigned a numerical code. For example, for the general health question, *"Would you say your health in general is excellent, very good, good, fair or poor?"*, two themes regarding participants' interpretation of the word *health* emerged: 1) a predominately physical conceptualization of health and 2) a multi-dimensional conceptualization including physical, but also mental, emotional and/or spiritual health. The following interview passages illustrate the two interpretive themes:

Physical Health: Coded 1 (physical):

"I'd say very good because I don't have any diseases, but I could be in better shape… you know, I should exercise more and it would be good if I stopped smoking completely."

Multi-dimensional Health: Coded 2 (multi-dimensional):

"My health is very good because I feel happy most of the time. I have a few aches and pains, but overall I feel good and I have a strong connection to God."

2 Results

Several types of analyses could be conducted using the final coded data set. Most broadly, problems were tallied to indicate a general frequency of problems posed by each question. To illustrate the wide range in problem severity as indicated by the overall measure used, Table 3 depicts (for 10 of the 30 total tested questions) the percentage of participants experiencing at least one type of coded problem (questions not illustrated revealed values that were intermediate in severity).

Table 3 Percentage of Subjects Producing at Least one Type of Question-Response Problem Code (for a Sample of Tested Questions)

Tested question	% subjects with 1+ codes
1. Do you have emphysema or chronic obstructive pulmonary disease (COPD)?	92.3% (60/65)
2. Do you have congestive heart failure?	77.5% (31/40)
3. What is the total value of all financial assets that you own? Please include Individual Retirement Accounts (IRAs), 401k plans, stocks and bonds, mutual funds, certificates of deposit (CDs), savings accounts, or any other financial assets.	78.1% (50/64)
4. Do you have chronic bronchitis?	53.7% (36/67)
5. Did you have a midday meal yesterday?	32.1% (17/53)
6. Do you have diabetes?	22.7% (15/66)
7. When you use butter or oils for cooking or preparing your food, which of the following types do you use most often? 1) Butter, Margarine, Lard, or Shortening, 2) Olive oil or Canola oil, 3) Corn oil, Vegetable oil, Peanut oil, Soy oil, 4) Non-stick spray, 5) Don't use fat	22.4% (15/67)
8. Would you say your health in general is excellent, very good, good, fair or poor?	11.9% (8/67)
9. Did you eat any other meals or snacks yesterday? (Other than the meals you just told me about)	4.5% (3/66)
10. Which fruit, vegetables, salad or juice did you have for a snack yesterday?	0.0% (0/66)

As illustrated in Table 3, a question on COPD (Chronic Obstructive Pulmonary Disease) scored the highest of the examined questions in this assessment. Understandably, all of the presented problems for this question were definition-based; 92% of the participants could not provide a correct definition for this term. Again, from this analysis it is not clear how serious this problem may be for estimate accuracy, as many participants were able to respond in a way that appeared to be accurate, based on further probing (however, it should be noted that a small proportion, thinking COPD equates to heart problems or asthma, answered the question incorrectly). However, this analysis did indicate that the problem itself is uncomplicated and could be easily fixed with a clarifying phrase or accompanying definition.

A question on total financial assets, on the other hand, presented a more complex set of problems. Only 1.6% of the problems associated with the asset question were definition-based. The primary problem pertained to inadequate response categories (61,0%), specifically that there was no adequate response category for those who had few or no such assets. Other problems involved question complexity (13.5%), inaccurate estimation (6.8%), lack of knowledge (8.5%) and question sensitivity (5,0%). Unlike the COPD question, problems involved every stage of the response process, and consequently there appeared to be no simple or straightforward modification that can be implemented to improve the question.

Statistical analysis of problem codes. Arguably, the types of analysis presented immediately above may not provide anything that a traditional cognitive analysis would omit, as any credible cognitive evaluation presumably should reveal these types of problems. The most advantageous aspects of this type of coding analysis, however, are (a) the ability to produce a quantitative estimate of potential problem severity (in terms of percentage of interviews in which a problem was in evidence), and (b) to explicitly investigate potential for group variation, by determining if a problem is more likely to occur within a particular group, as opposed to evenly distributed across respondents. To determine if problems were systematically related to ethnicity or to other measured subject characteristics, cross-tabulations and logistic regression analyses were conducted, involving age, gender, ethnicity, income, and education. Logistic regression analysis was done in hierarchical manner, with ethnicity (Hispanic/Non-Hispanic) entered last into the model, so as to determine the unique contribution of this variable, controlling for the influence of other measured demographics. Because 12 items exhibited either ceiling or floor effects (defined as >90% interviews illustrating error, and <10%, respectively), 18 items were considered statistically analyzable.

The regression analysis determined that for these 18 items, ethnic group membership was the strongest overall predictor of problem code frequency ($p < .05$), with Hispanics generally experiencing more difficulties than Non-Hispanics (for 5 items), but with Hispanics seemingly having fewer problems for two other evaluated questions. Somewhat surprisingly, normally analyzed demographic characteristics, including gender, age, educational level, and income, had relatively weak effects: Gender produced no significant effects for any item; older respondents had somewhat more trouble than younger ones with a question on lifting and carrying; and subjects with lower income produced more codes for a question concerning coronary heart disease. Subjects with lower educational level produced significantly more codes for the question *"Do you have chronic bronchitis?"*: 75% of those participants who did not graduate from high school, as opposed to 30% of those with a high school degree, exhibited some type of problem with this question. An exami-

nation of the qualitative interviewing results reveals that the difficulty centered primarily around uncertainty of the word *chronic*; in particular, participants with lesser education confused episodes of *acute* and *chronic* bronchitis. As a follow-up analysis, 80% of the more educated participants (along with 100% of the less educated) were found to have experienced terminology problems in the chronic obstructive pulmonary disease question, suggesting that few individuals are familiar with terminology associated with chronic disease that is commonly used within health surveys.

Overall, effects as measured by summary problem codes were mainly related to ethnic group membership. Even these were not consistently unidirectional, however. Hispanics produced significantly more problems with a question on ever having cancer, and significantly less for one on combined household income. Five other questions exhibiting Hispanic/Non-Hispanic differences involved food and meal questions; data are depicted in Table 4.

Table 4 Percentage of Participants Having Response Problems, by Ethnicity

Tested question	Hispanics	Non-Hispanics
1) How many times did you eat red meat, including beef, pork, lamb, or lunchmeat, hot dogs or sausages made from beef, pork or lamb yesterday?	77.1% (27/35)	35.5% (11/31)
2) Did you eat a morning meal yesterday?	66.7% (24/36)	6.5% (2/31)
3) Did you eat a midday meal yesterday?	54.2% (13/24)	13.8% (4/29)
4) Did you eat an evening meal yesterday?	57.7% (15/26)	6.7% (2/30)
5) When you use butter or oils for cooking or preparing your food, which of the following types do you use most often? 1) Butter, Margarine, Lard or Shortening, 2) Olive oil or Canola oil, 3) Corn oil, Vegetable oil, Peanut oil, Soy oil, 4) non-stick spray, 5) Don't use fat	8.3% (3/36)	38.7% (12/31)

Again, qualitative analysis facilitates interpretation of these differences, and to pinpoint the character of the response problem. From the very beginning of Spanish language interviewing, it was clear that some translated survey questions caused interpretation difficulties for Hispanic subjects. That is, particular words were translated literally from English and, because of cultural differences, did not convey the same meaning. For example, the phrase *frijoles con chile* was intended to mean chili beans, but was interpreted by most Hispanic participants as beans with hot sauce. Additionally, some words varied by particular region (e.g., Puerto Rican Spanish uses *nami* for yam, while Mexican Spanish

uses *camote*) or were more formal forms of Spanish (e.g., the word *fiambre* for lunch-meat). Consequently, these terms were not always understood by Hispanic participants. It is this variety of translation problem that apparently accounted for the higher percentage of Hispanics experiencing problems with the red meat question.

Similarly, some words in Spanish consisted of more than one meaning and could easily be taken out of context. For example, the word *comida* can mean *meal, food*, and the name of a meal – like the English word for dinner. Consequently, the question *"Did you eat a morning meal?"* was translated as *"¿Ayer comío Ud. la comida de la mañana?"* but misunderstood by some Hispanic participants as *"Did you eat your dinner in the morning?"* This interpretive and translative issue accounts for a large portion of the ethnicity-based response problems regarding the meal questions, and is evidenced by the following exchange, which also illustrates how cognitive probing brings out an otherwise "silent misunderstanding," as termed by DeMaio & Rothgeb (1996):

Interviewer:	*Digame, Ayer comio usted la comida de la manana?*
	Tell me, did you eat a meal in the morning?
Participant:	No.
Interviewer:	*Y la manana para Usted, que quiere decir, que tanto tiempo, de que horas a que horas?*
	And morning for you, what does it mean, what time frame or from what hour to what hour is it?
Participant:	*Pues en la manana el desayuno es a las nueve.*
	Well in the morning el desayuno is at nine.
Interviewer:	*So, el desayuno, lo nombra el desayuno, es a las nueve?*
	So, el desayuno, you name it desayuno, is at nine?
Participant:	*Si, por que you no doy el que le dicen... .como le dicen... Braaq faat*
	Yes, because I don't serve, what they call… how do they say…Braaq faat
Interviewer:	Breakfast?
Participant:	*Si.*
	Yes.
Interviewer:	*No hace breakfast sino que hace desayuno?*
	You don't make breakfast, but you make desayuno?
Participant:	*Si, yo desayuno, asi estoy acustumbrada…doy mi desayuno y mi comida y en la cena como algo mas liviano.*
	Yes, I have desayuno, that is how I am accustomed… I serve desayuno, and then my comida and for cena, I eat something a lot lighter.

In this case the subject was initially asked about "comida de la manana," or as intended by the investigators, "food in the morning." However, she evidently interprets "comida" as the meal eaten after her morning meal (which is "desayuno"), and so indicates that no, she has not eaten her "comida" meal in the morning – and produces what, to the survey administrator, is an erroneous response.

A qualitative examination of the interviews also reveals why Non-Hispanic subjects, in comparison to Hispanics, were likely to experience problems with the oil questions. Many Hispanics reported using butter and lard to cook, and, consequently were able to provide an answer with little consideration. Non-Hispanic participants, on the other hand, were much more likely to use a variety of cooking oils and experienced trouble determining which type of oil they used most often. To provide an answer, they often needed to mentally recreate their cooking habits – which oil they used for which specific type of food – and then consider which oil was used most often.

Analysis of the interpretive codes, while not necessarily signaling potential response error, also illustrates how ethnicity impacts the question-response process, particularly question comprehension. For example, from examination of the problem codes, there appeared to be few problems in the general health question, *"Would you say your health in general is excellent, very good, good, fair or poor?"* Those problems that were identified were related to response categories – in which participants had difficulty matching their personal conceptualization of health within the provided categories. However, analysis of the interpretive codes reveal an intriguing finding: While most of the Non-Hispanic participants (77%) gave reports coded as conceptualizing health as a physical phenomena, most of the Hispanic participants (90%) used a comprehensive conceptualization of health, incorporating emotional and spiritual dimensions. It is not surprising that Hispanic participants – especially those who were female, and those raised in Mexico – more closely associated health with spirituality, as traditional Mexican medicine, or *curanderismo*, is directly connected with ritual and a more holistic sense of well-being. It is interesting to note that the few Hispanic participants who did not hold a comprehensive view of health were second generation Mexican Americans, and consequently, may have assumed Non-Hispanic cultural customs. At this point, the extent to which differing conceptions of health (as well as other culturally-based interpretive differences) may impact the quality of survey data is unclear. However, it is evident that because of varying cultural interpretations (of even a single term), participants were in effect answering two distinctly different questions.

Finally, some obtained results were unrelated to language or culture. During the course of conducting the interviews, it became obvious that, for the question *"Yesterday did you eat*

any beans such as kidney beans, refried beans, chili beans, bean soup, bean salad or lentils?", participants adopted differing interpretations of the word *bean*. Some participants viewed the question as asking about legumes only, while others included any kind of bean, even green beans. Yet, at that point, interviewers could only speculate whether there was a particular group of participants using a specific interpretive pattern. It was hypothesized that older participants and perhaps less educated participants would be less inclined to view the question as asking about legume consumption. As it turned out, based on statistical (regression) analysis, the patterns of interpretation were not related to specific demographic group membership, as "green bean error" was found to be essentially random.

3 Discussion

This paper has attempted to show how various types of analyses can be conducted using coded, semi-structured cognitive interviews which examine relationships between participant characteristics, such as ethnicity, and responses to tested survey questions. Further, based on the quantification of these codes, it may be determined that certain problems are more serious than others. The nature and causes of these problems can then be further investigated by making use of the original, qualitative interviewer notes from which the codes were developed. From the current study, it does appear that non-trivial differences may exist between Hispanics and Non-Hispanics in answering common health survey questions. Some of these are due to translation, some to cultural influences; some favor Non-Hispanics, others Hispanics. Presumably, problems with the translated version can be addressed through revisiting the translation process, or avoided in the first place by applying effective translation techniques (see Harkness, van de Vijver & Mohler, 2003; McKay, Breslow, Sangster, et al., 1996). Problems that have a basic cultural origin that transcends language may require more careful consideration of whether the questions as posed apply equally to all major groups to be included in the survey, and whether some underlying assumptions must be revisited (Ainsworth, 2000). In all cases, the mixed-method approach appeared to have significant utility.

Limitations and caveats. Several limitations to the existing study, and potential weaknesses of the evaluated methodological approach, must also be acknowledged:

(1) The practice of counting up problems within a small sample of cognitive interview subjects can be misleading, as this does not necessarily create a reliable index of problems that will then occur within in a survey field environment (Willis, 2005). Some problems obtained in a single cognitive interview may be of critical importance, to the extent that

these may in turn predict serious difficulty or source of error in the field environment for an important segment of the surveyed population.

(2) Further compounding this problem, in many pretesting studies, much smaller rounds of interviews are generally conducted than were included in the current study; restrictions by the U.S. Office of Management and Budget often limit the size of an interviewing round to no more than nine subjects. In such cases it is unlikely that strict quantification of results would supply the amount of information obtained when many more interviews are conducted, and the quantifiable arm of the mixed-method approach may be woefully insufficient.

(3) Because Hispanic/Spanish language interviews were conducted in Ohio but not in Washington D.C., the current study partially confounded ethnic group membership and region in which the interview was done; as Miller (2002) has observed, cognitive interviews that are conducted in varying regions can produce somewhat different results, and lead to varied conclusions concerning the adequacy of individual survey questions. Hence, Hispanic versus Non-Hispanic differences observed conceivably could, to some extent, reflected regional rather than ethnic or language variation[1].

(4) By far, the greatest drawback experienced in the current study was the amount of time required to ensure that all codes were applied consistently across all interviews, and to develop and clean an analyzable data set. The current approach cannot practically be implemented within cognitive testing projects that must be completed quickly or with little cost.

On the other hand, it was found that use of a mixed method approach was advantageous in providing multiple types of findings that are not entirely obtainable with traditional analytic methods for assessing cognitive interviews. Because both quantitative and qualitative analyses were used, not only could a rough measure of problem intensity for each tested question be obtained, but also the nature of potential response errors, and whether the problem appeared to be systematically related to demographic or other measured characteristics. Unlike behavior coding (Fowler & Cannell, 1996), mixed-method coding is intensively contextual in nature and allows for a more in-depth understanding of problem origins. Perhaps most importantly, this approach provides another avenue for theory building within the field of question design. For example, examining questions that are influenced by demographic characteristics will provide a better understanding of the

1 On the other hand, note that demographic variables that are commonly associated with regional differences, such as educational level and income, were fairly well controlled in the current study, so there is no particular reason to suspect a region effect in this case.

relationship between respondents' social location and response error and, in turn, will provide insight into the quality of survey estimates for particular socio-cultural groups. Currently, a multi-agency effort (involving NCHS, the U.S. Census Bureau, Bureau of Labor Statistics, and National Institutes of Health) is underway to apply such a scheme routinely to the outcomes of cognitive interviewing projects, in order to produce a dataset useful for establishing general relationships between question characteristics, respondent characteristics, and sources of error in survey questions (Miller, Canfield, Beatty, et al., 2003).

References

Ainsworth, B. E. 2000. "Issues in the assessment of physical activity in women." *Research Quarterly for Exercise and Sport* 71:37-42.

Conrad, F., J. Blair, and E. Tracy. 2000. *Verbal reports are data! A theoretical approach to cognitive interviews.* Office of Management and Budget: Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference, 317-326.

DeMaio, T. J., and J. M. Rothgeb, 1996. "Cognitive interviewing techniques: In the lab and in the field." Pp. 177-195 in *Answering questions: Methodology for determining cognitive and communicative processes in survey research,* edited by Schwarz, N., and S. Sudman. San Francisco: Jossey-Bass.

Denzin, N., and Y. Lincoln. 2000. *The Handbook of Qualitative Methods.* Thousand Oaks: Sage Publications.

Foddy, W. 1998. "An empirical evaluation of in-depth probes used to pretest survey questions." *Sociological Methods and Research* 27:103-133.

Fowler, F. J., & C. F. Cannell. 1996. "Using behavioral coding to identify problems with survey questions." Pp. 15-36 in *Answering questions: Methodology for determining cognitive and communicative processes in survey research,* edited by Schwarz, N., and S. Sudman. San Francisco: Jossey-Bass.

Gerber, E. 1999. "The view from anthropology: Ethnography and the cognitive interview." Pp. 217-234 in *Cognition and Survey Research,* edited by Sirken, M. G. et al. New York: Wiley.

Harkness, J. A., F. J. R. van de Vijver, and P. Mohler. 2003. *Cross-Cultural Survey Methods. Hoboken.* New York: Wiley.

Johnson, T. P. 1998. "Approaches to equivalence in cross-cultural and cross-national survey research." *ZUMA-Nachrichten Spezial* 3:1-39.

McKay, R. B., M. J. Breslow, R. L. Sangster, S. M. Gabbard, R. W. Reynolds, J. M. Nakamoto, and J. Tarnai. 1996. "Translating survey questionnaires: Lessons learned." *New Directions for Evaluation* 70:93-105.

Miller, K. 2002. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St Pete Beach, FL. *The role of social location in question response: Rural poor experience answering general health questions.*

Miller, K., B. Canfield, P. Beatty, K. Whitaker, and B. Wilson. 2003. Paper presented at the 2003 Federal Committee on Statistical Methodology Research Conference, Arlington, VA. *Q-BANK: Development and implementation of an evaluated-question database.*

Redline, C., R. Smiley, M. Lee, T. DeMaio, and D. Dillman. 1998. *Beyond concurrent interviews: An evaluation of cognitive interviewing techniques for self-administered questionnaires* (Working Papers in Survey Methodology No. 98/06). Washington, DC: U.S. Census Bureau. Retrieved February 12, 2004 from www.census.gov/srd/papers/pdf/sm98–06.pdf.

Snijkers, G. 2002. *Cognitive laboratory experiences. On pretesting computerised questionnaires and data quality.* Unpublished Doctoral Dissertation, Utrecht University, Netherlands.

Snijkers, G. 2004. *Cognitive laboratory experiences and beyond: Some ideas for future research.* Proceedings of the 4th Conference on Questionnaire Evaluation Standards, Zentrum fur Umfragen, Methoden und Analysen (ZUMA), 190-203.

Tashakkori, A., and C. Teddlie. 1998. *Mixed Methodology: Combining Qualitative and Quantitative Approaches.* Thousand Oaks: Sage Publications.

Tourangeau, R. 1984. "Cognitive science and survey methods: A cognitive perspective." Pp. 73-100 in *Cognitive aspects of survey design: Building a bridge between disciplines,* edited by Jabine, T. et al. Washington, DC: National Academy Press.

Tucker, C. 1997. "Measurement issues surrounding the use of cognitive methods in survey research." *Bulletin de Methodologie Sociologique* 55:67-92.

Warnecke, R. B., T. P. Johnson, N. Chavez, S. Sudman, D. P. O'Rourke, L. Lacey, et al. 1997. "Improving question wording in surveys of culturally diverse populations." *Annals of Epidemiology,* 7(5):334–342.

Willis, G. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design.* Thousand Oaks: Sage.